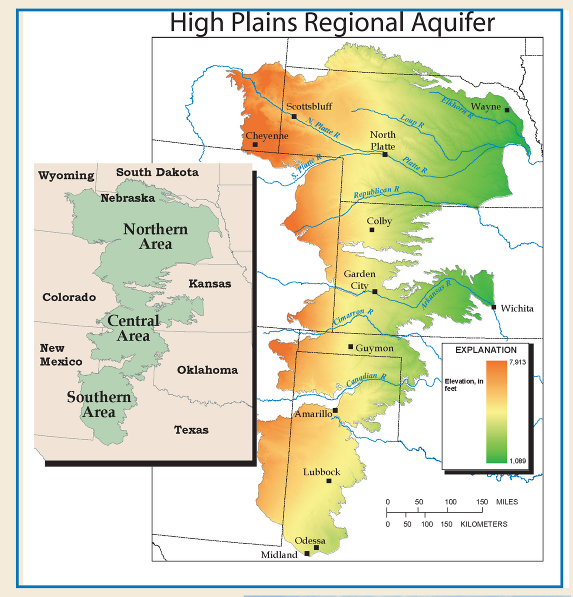
The Use of GIS in Modeling Ground-Water Vulnerability to Nitrate in the High Plains aquifer

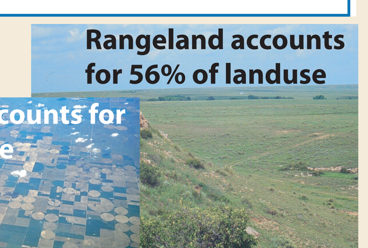
Sharon L. Qi

Jason J. Gurdak

Abstract

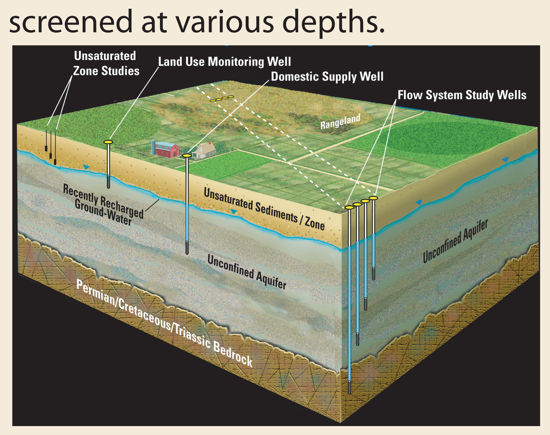
GIS is often used as an important tool in developing ground-water vulnerability models and corresponding maps, which are valuable for ground-water resource management and planning. In this study, which is an expansion of a pilot study, ArcGIS Desktop and Workstation were used to extract geospatial data from various large data sets for input to a logistic regression model of ground-water vulnerability and to produce a corresponding vulnerability map. The map illustrates the predicted probability of recently recharged (defined as less than 50 years) ground water of the High Plains aquifer to non-point source nitrate contamination. Spatial data from 31 individual vector and raster layers were extracted for each of 6,946 well locations throughout the study area. These layers included information about saturated thickness, depth to water, precipitation, percent irrigated/nonirrigated/agricultural land, nitrogen applications, soil characteristics, lithology of unsaturated zone, playa lakes, and water use. Extractions for categorical data and certain continuous data sets were performed using a series of identity overlays, directly from the layer at the location of each well. Ninety-degree sectors of varying radii, which was determined by hydraulic conductivity, were created upgradient from each well for extraction of data where information needed to be related to a well location. These variables were inventoried for the 90-degree sector areas using both vector-union techniques and raster map-algebra techniques. The extracted data were used as input for logistic regression analyses to determine which of the variables (layers) or combination of variables were significantly correlated with observed water-quality conditions and would be used in a model of the probability of predicting nitrate concentrations greater than 4 milligrams per liter (as N) in ground water. Five variables were considered significant (depth to water, organic content of soils, amount of irrigated/nonirrigated land, and the amount of clay in the unsaturated zone) in the final two models that were developed to represent different regions of the study area. The appropriate GIS layers were converted to raster data sets in order to use the map algebra capabilities of ArcGIS. The two model equations and the various coefficients for each layer were fed back into ArcGIS and, using map algebra, the probability surface was calculated and then easily visualized across the entire study area with GIS.





GIS Extraction of Explanatory Variables

Select from > 8,000 wells with NO_3 data, screened at various depths.

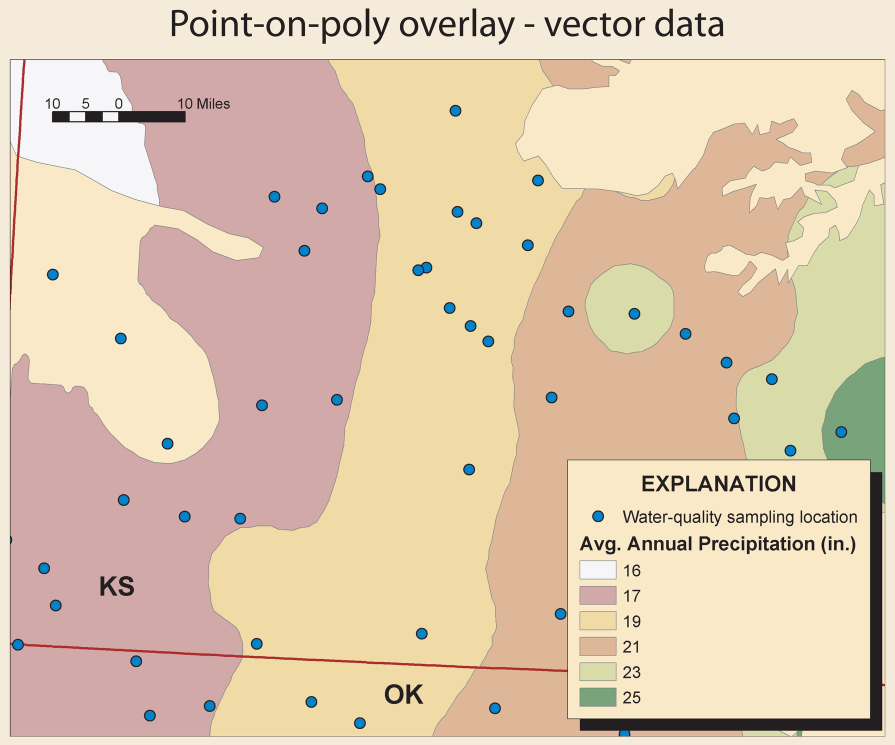


Spatial data were extracted in two ways. For categorical data, the information was extracted from the layer at the well point using identity overlays. For other continuous data sets where the information needed to be summarized for each well, ninety degree-sector buffers were created and the data were summarized using both vector union techniques and raster map algebra techniques.

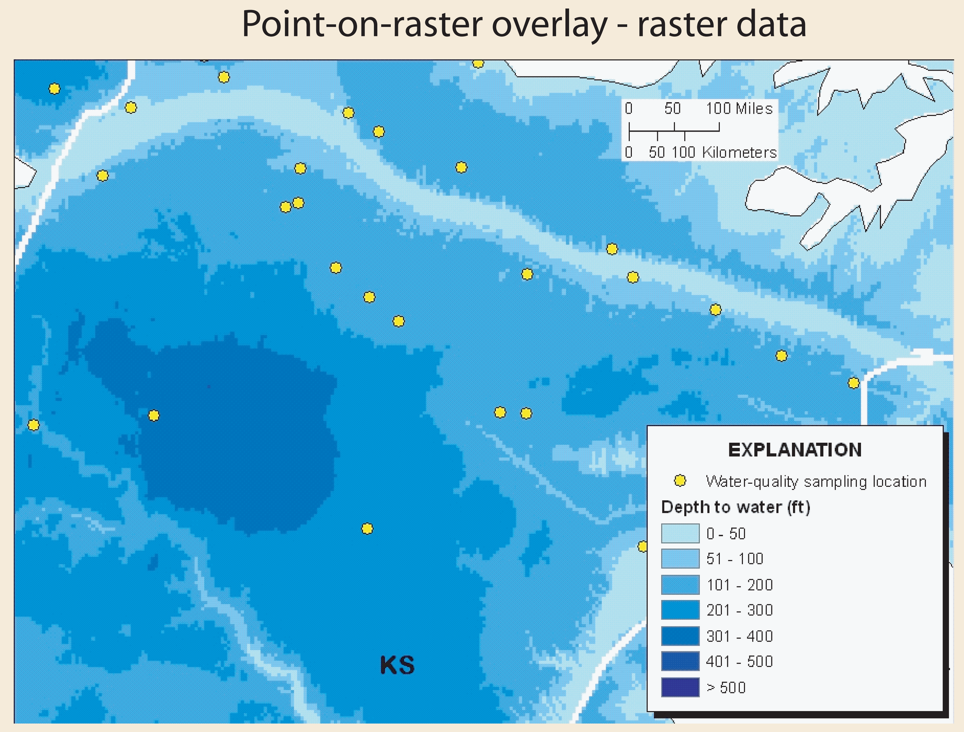
The data for each well were then exported as tables for import into the statistical package where the models were created.

For Catagorical data and certain raster data sets (large grid-cell size), data extraction was by simple point overlay:

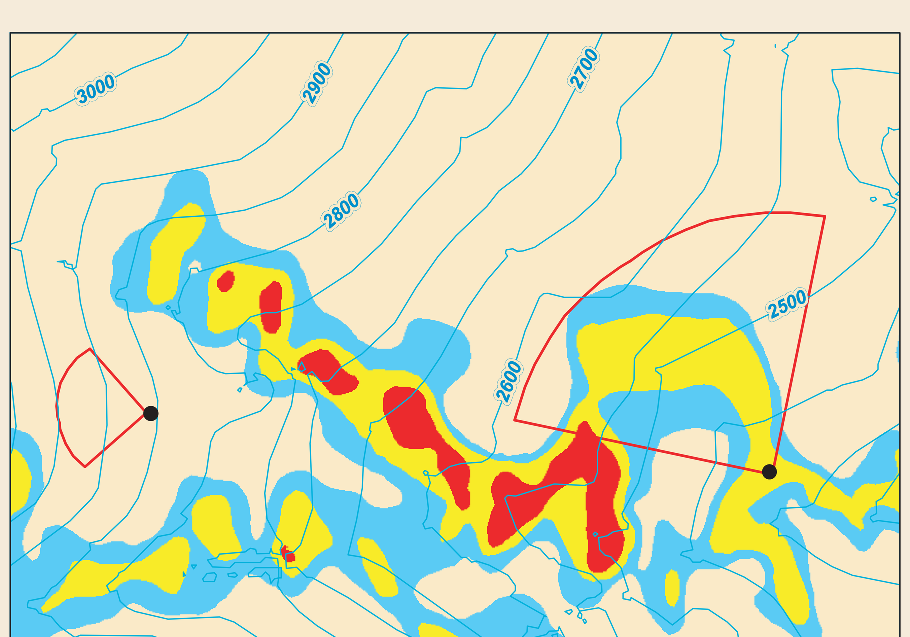
Point-on-poly overlay - vector data



Point-on-raster overlay - raster data

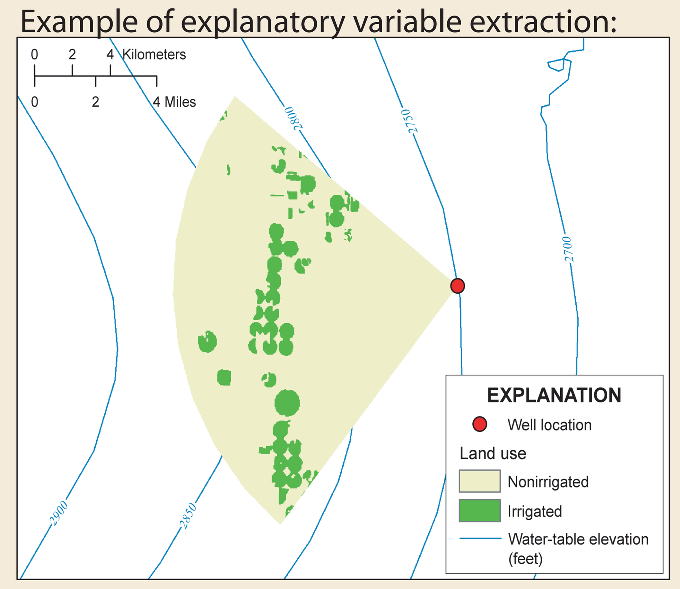


For selected continuous data sets, data extraction was done using ninety degree-sectors. Results of horizontal particle-tracking defined length of contributing area, based on estimated hydraulic conductivity:



Contributing areas, delineated by a 90-degree sector, represent the capture zone up-gradient from the well location and were used to define the area for GIS-based extraction of explanatory variables for statistical modeling.

Example of explanatory variable extraction:

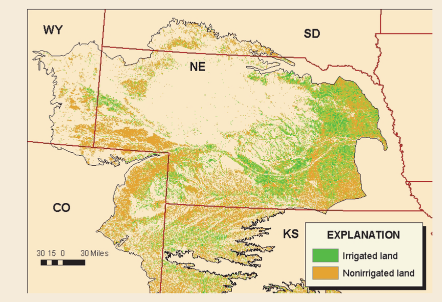


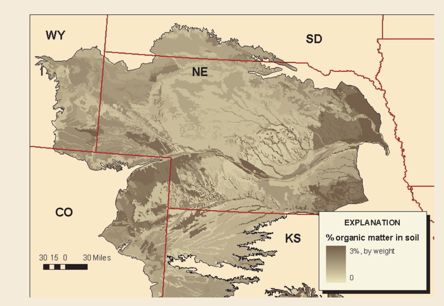
For additional information: http://webserver.cr.usgs.gov/nawqa/hpgw/index.html

Two Models

Final multivariate logistic regression models were selected based on statistical significance, model fit, and predictive ability; one for the northern High Plains and one for the combined central and southern High Plains:

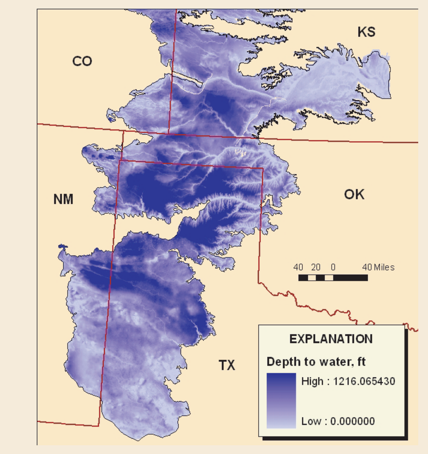
Northern High Plains: significant variables are amount of irrigated/nonirrigated land upgradient of well and amount of organic matter in soil.

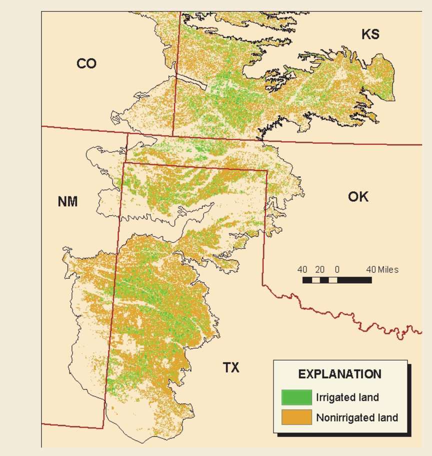


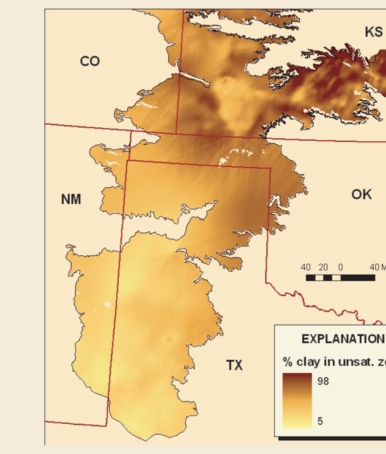


$$\text{Probability (NHP)} = \frac{e^{(-0.374 + 0.023*\text{nonirrigated land} + 0.017*\text{irrigated land} + -1.487*\text{soil org. matter})}}{1 + e^{(-0.374 + 0.023*\text{nonirrigated land} + 0.017*\text{irrigated land} + -1.487*\text{soil org. matter})}}$$

Central and southern High Plains: significant variables are depth to water, amount of irrigated/nonirrigated land upgradient of well and amount of clay in the unsaturated zone upgradient of a well.







$$\text{Probability (NHP)} = \frac{e^{(1.158 + -0.010*\text{DTW} + 0.043*\text{nonirrigated land} + 0.011*\text{irrigated land} + -0.019*\text{pct clay in unsat. zone})}}{1 + e^{(1.158 + -0.010*\text{DTW} + 0.043*\text{nonirrigated land} + 0.011*\text{irrigated land} + -0.019*\text{pct clay in unsat. zone})}}$$

Uncertainty Analysis

Addresses the confidence surrounding the estimates of ground-water vulnerability:

- Latin Hypercube (stochastic) sampling provided estimates of model uncertainty due to:
  - Model error: confidence interval of the logistic regression model coefficients
  - Data error: estimated spatial uncertainty due to explanatory variable value and position
- 95% prediction intervals were estimated and represented in GIS, and identify smaller uncertainty associated with extreme probabilities and larger uncertainty associated with predicted probabilities approximating 50%

Relative Variance Contribution (RVC_r)

$RVC_r = (\sigma_r / \sigma) \times 100\%$

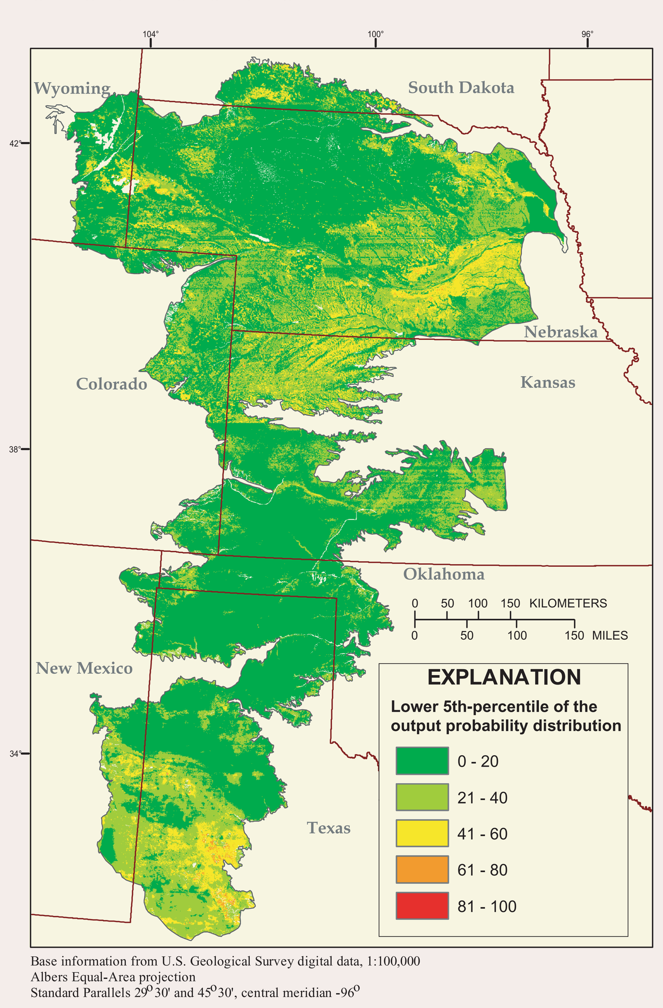
Where: σ_r = regression coeff. variance

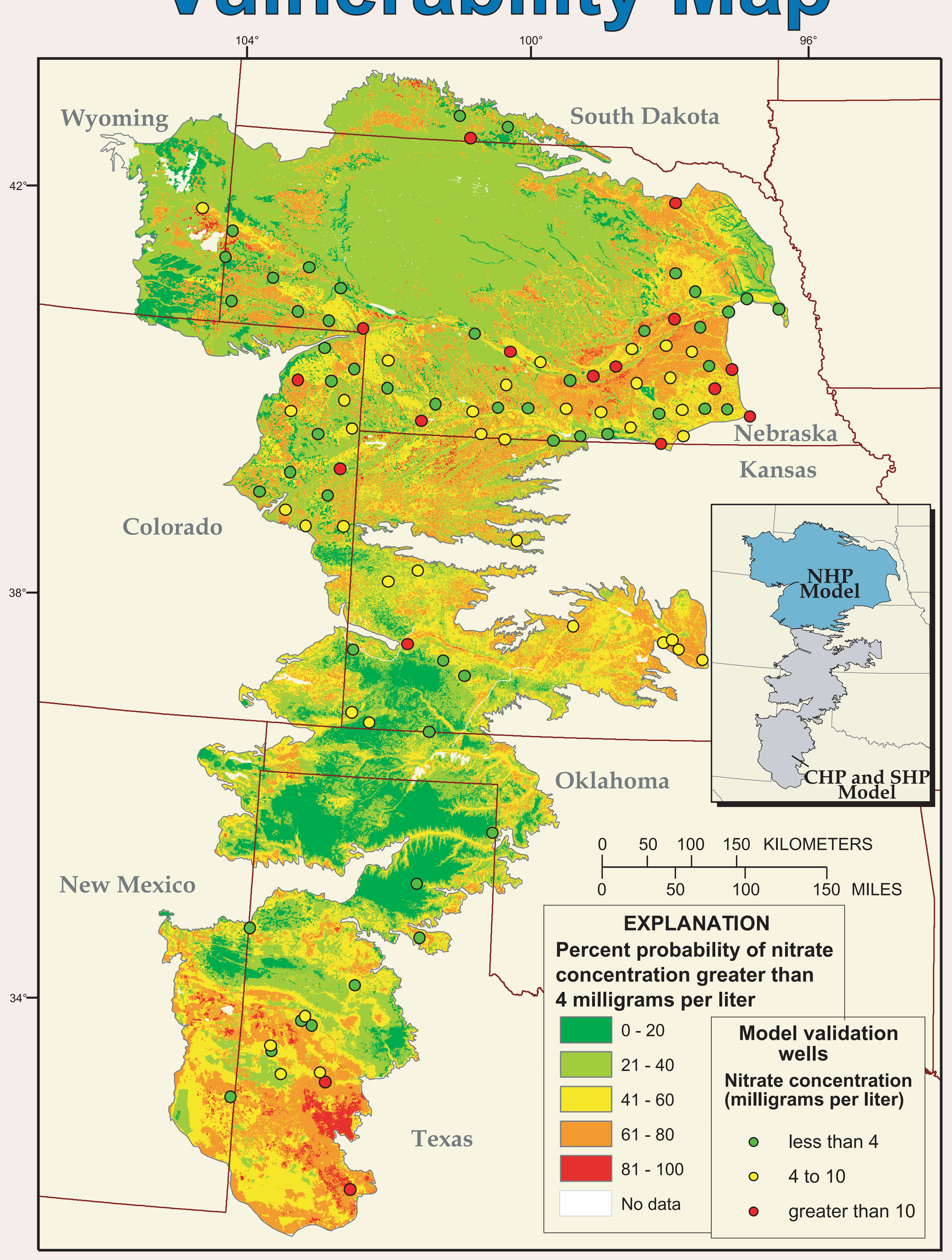
σ = variance from uncertainty in explanatory variables

Model has $RVC_r > 50\%$, indicating uncertainty due to regression coeff. dominates the total uncertainty

Vulnerability Map

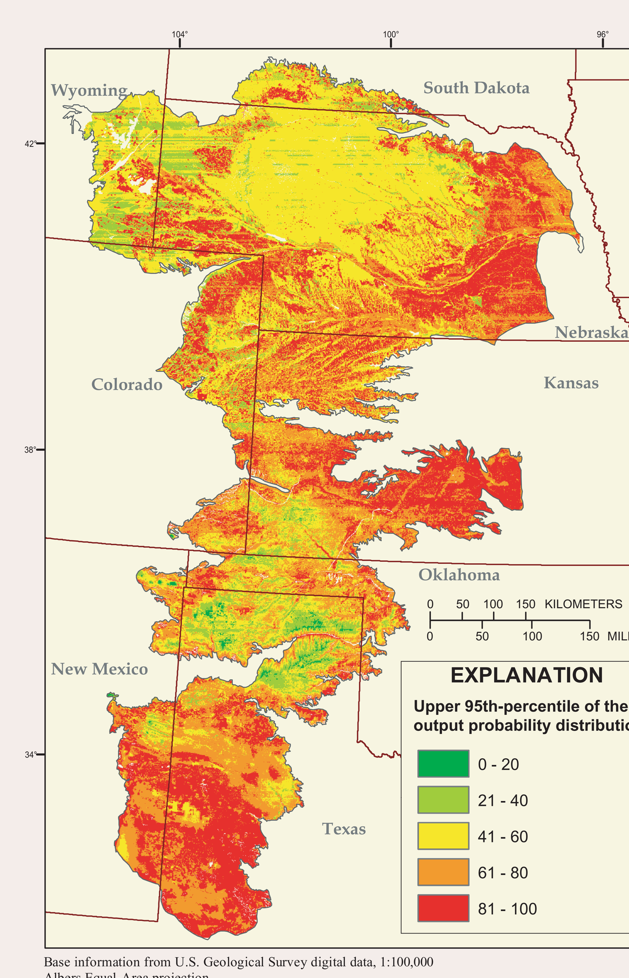
Lower 5th-percentile of output probability distribution





Vulnerability map illustrates the predicted probability of detecting nitrate > 4 mg/L in recently recharged groundwater (< 50 years). Nitrate concentrations from validation well network is shown.

Upper 95th-percentile of output probability distribution



Conclusions

- The vulnerability model and map offer a predictive tool for water resource managers to identify likely areas of non-point source contamination and evaluate the impact of anthropogenic activity on nitrate distribution in groundwater.
- The final vulnerability model and map of the High Plains aquifer corroborated our conceptual model that nitrate concentrations are directly related to nitrogen loading and transport mechanisms at land surface and infiltration in the soil zone, and inversely related to impedances to downward advective chemical movement through the unsaturated zone.
- Uncertainty analysis is important because it provides an estimate of the magnitude of error associated with the prediction of ground-water vulnerability, and would be of value for resource managers required to make best informed decisions.
- Additional nitrate samples are needed in areas of mid-range predicted probabilities to reduce model uncertainty and increase sensitivity.